Statistical Tests of Salmon Fishery Contribution Rates Based on Coded-Wire-Tag Recoveries

Ken Newman[1] and Richard Comstock

U.S. Fish and Wildlife Service

2625 Parkmont Lane SW, Bldg A

Olympia, WA 98502

[1]and Department of Statistics, University of Washington, Seattle, WA 98195

Abstract

Statistical procedures for testing hypotheses concerning fishery contribution rates based on recoveries of coded-wire-tagged (CWT) salmon are described. Observed recoveries of CWT fish from one or more release groups to one or more fisheries are modeled with compound probability distributions, one component reflecting natural variation and the other representing the catch sampling variation. We describe a flexible bootstrap testing procedure suitable for a wide range of hypotheses, verify its accuracy by simulation of compound probability distributions, and apply it to a set of real CWT recoveries.

1 Introduction

Fishery managers and scientists raise many questions about the survival, the catch distribution, and the migration patterns of salmon. Some of these questions can be formulated in terms of contribution rates to one or more fisheries (including escapement). For example, does a particular feeding regime at a hatchery increase the contribution rate to a specified fishery? Or, more generally, does a set of hatchery treatments have differing effects on a vector of contribution rates? Do the hatchery releases on a river system have the same catch distribution or profile as the wild stock native to the river system?

Recoveries of coded-wire-tagged (CWT) salmon in fisheries' catches and on the spawning grounds provide the basic data for attempting to answer these questions. The primary objective of this report is to describe a flexible procedure for testing some of these hypotheses using information provided by CWT recoveries.

Coded-wire-tags are implanted in the heads of young salmon, generally at the smolt stage, that are about to be released from a hatchery or that are wild stock netted in a river system. Later in the life cycle the catches of intercepting fisheries are subsampled to recover the adult CWT fish. A compound probability distribution, the binomial-hypergeometric, to model the number of recoveries of CWT salmon in a fishery's catch was proposed earlier by Newman (1990). (A mathematically similar, but conceptually different, model was proposed by Clark and Bernard (1987).) This probability distribution explicitly accounts for two different, successive levels of randomness or uncertainty that induce variation in the number of observed recoveries in a catch sample. Conventional testing approaches such as t-, chi-square, and F tests are inappropriate under the compound distribution model because of the second level of variation induced by catch sampling. In this paper we present a general statistical procedure for conducting hypothesis tests

3

based on compound probability distributions.

The structure of the paper is as follows. First we briefly describe the binomial-hypergeometric distribution and present some extensions. Next two hypotheses of general interest are briefly described. The following section describes monte carlo testing, bootstrap testing, and the iterated bootstrap testing procedure proposed by Beran (1988). The next section presents both the results of computer simulations as well as results for a real data set. The last section discusses the procedure and suggests areas of additional work.

2   Compound distributions for CWT recoveries

2.1   Notation and terminology

The following symbols will be used.

p_{ij}: contribution rate or survival rate of release group i to fishery j

r_{ij}: relative contribution rate to fishery j of release group i, i.e., the contribution of group i to fishery j divided by group i's contribution to all fisheries

R_i: release group i's size

a_{ij}: actual contribution of release group i to a fishery j

f_j: catch sampling fraction for fishery j

o_{ij}: observed recoveries of a release group i in fishery j's catch sample

b_j: contribution of fish other than relevant release groups to fishery j

$C_j = \sum_i a_i + b$, total catch of fishery j

$n_j = f_j \cdot C_j$. catch j sample size (in practice an integer)

Contribution rates of interest are often the sums of contribution rates to several sub-fisheries. By sub-fishery is meant a fishery for which a single sample of the catch is drawn. For example. the contribution rate of coho salmon. *Oncorhynchus*

kisutch, from a specific release group to the Oregon ocean troll fishery in June is the sum of contribution rates made to all the different sub-areas off the Oregon coast by day and landing.

2.2 The distribution of observed recoveries

The assumptions are as follow. A release group contains R tagged fish, the fate of each fish with respect to its being caught or not caught by a specific fishery, designated by a 1 or 0, respectively, is independently and identically distributed as a Bernoulli(p) random variable. A simple random sample is drawn without replacement from the fishery's catch, which includes fish from other sources as well. The catch sampling rate is a fixed rate, and the catch of other fish is also assumed to be a fixed number.

From these assumptions a compound distribution results. The distribution of the number of tagged recoveries in the catch, a, is Binomial(R,p). Conditioning on the number of tagged fish in the total catch, the number of tagged fish recovered in the sample, o, follows a hypergeometric distribution, $o \sim$ Hypergeometric(C, a, n). The distribution of observed fish is found by summing the joint distribution of tagged fish caught in the fishery and tagged fish observed in the catch sample over all possible values of tagged fish in the fishery:

$$\Pr[O = o] = \sum_a \Pr[a, O = o] = \sum_{a=o}^{R} \binom{R}{a} p^a (1-p)^{R-a} \cdot \frac{\binom{a}{o}\binom{b}{n-o}}{\binom{a+b}{n}}.$$

This distribution is referred to as the binomial-hypergeometric distribution for observed recoveries.

If the contribution of two release groups of sizes R_1 and R_2, respectively, to a single fishery are considered, the joint distribution of the observed recoveries is product binomial-multiple hypergeometric:

$$\Pr[O_1 = o_1, O_2 = o_2] = \sum_{a_1} \sum_{a_2} \Pr[a_1, a_2, O_1 = o_1, O_2 = o_2] =$$

$$\sum_{a_1=o_1}^{R_1} \sum_{a_2=o_2}^{R_2} \binom{R_1}{a_1} p_1^{a_1} (1-p_1)^{R_1-a_1} \cdot \binom{R_2}{a_2} p_2^{a_2} (1-p_2)^{R_2-a_2} \cdot \frac{\binom{a_1}{o_1} \binom{a_2}{o_2} \binom{b}{n-o_1-o_2}}{\binom{a_1+a_2+b}{n}}.$$

The 'product' binomial reflects the independence of the two release groups.

If two release groups of sizes R_1 and R_2, respectively, contribute to two fisheries, say, then the joint distribution of observed recoveries in the two fisheries from the two release groups is product multinomial-multiple hypergeometric:

$$\Pr[O_{11} = o_{11}, O_{12} = o_{12}, O_{21} = o_{21}, O_{22} = o_{22}] =$$

$$\sum_{a_{11}=o_{11}}^{R_1} \sum_{a_{12}=o_{12}}^{R_1-a_{11}} \sum_{a_{21}=o_{21}}^{R_2} \sum_{a_{22}=o_{22}}^{R_2-a_{21}} \Pr[a_{11}, a_{12}, a_{21}, a_{22}, O_{11} = o_{11}, O_{12} = o_{12}, O_{21} = o_{21}, O_{22} = o_{22}] =$$

$$\sum_{a_{11}=o_{11}}^{R_1} \sum_{a_{12}=o_{12}}^{R_1-a_{11}} \sum_{a_{21}=o_{21}}^{R_2} \sum_{a_{22}=o_{22}}^{R_2-a_{21}} \frac{R_1!}{a_{11}! a_{12}! (R_1 - a_{11} - a_{12})!} \frac{R_2!}{a_{21}! a_{22}! (R_2 - a_{21} - a_{22})!} \cdot$$

$$p_{11}^{a_{11}} p_{12}^{a_{12}} (1-p_{11}-p_{12})^{R_1-a_{11}-a_{12}} \cdot p_{21}^{a_{21}} p_{22}^{a_{22}} (1-p_{21}-p_{22})^{R_2-a_{21}-a_{22}} \cdot \binom{a_{11}}{o_{11}} \binom{a_{21}}{o_{21}} \binom{b_1}{n_1-o_{11}-o_{21}} \binom{a_{12}}{o_{12}} \binom{a_{22}}{o_{22}} \binom{b_2}{n_2-o_{12}-o_{22}}$$

One assumption that is clearly false is that the catch of other fish, b, is a constant. However, results of computer simulations treating catch as a random variable or as a constant have not indicated that this is a critical assumption in terms of variance estimation of the estimated contribution rate or in hypothesis testing.

More critical, perhaps, is the assumption of independence at the Bernoulli stage given the fact that fish school. This issue was studied for various ocean troll fisheries and there was no evidence for clustering of the recoveries from the same release group. More work is necessary, however, to determine if the lack of clustering was due more to randomness in the troll fishing process than actual independence in the fish's behavior.

2.3 Estimation of contribution rate

The estimator of contribution rate used in practice is quite simple and natural. If the total number of tagged fish in the catch were known, namely a, then $\hat{p} = a/R$ would be the standard estimator for a Bernoulli sample. However, since a is unknown and an unbiased estimate of a is $\hat{a} = o \cdot C/n$, a natural estimate of p is $\hat{p} = \frac{o \cdot C}{n \cdot R} = \frac{o}{f \cdot R}$.

This is a method of moments estimator since the catch sample proportion of tagged fish is substituted for the true proportion of tagged fish and the estimated true proportion of tagged fish is substituted for the distributional proportion. It is also unbiased:

$$E(\hat{p}) = \frac{1}{f \cdot R} \cdot E(o) = \frac{1}{f \cdot R} \cdot E_a E(o|a) = \frac{1}{f \cdot R} \cdot E_a(f \cdot a) = p.$$

Method of moments estimators are asymptotically consistent, i.e., as the sample size, or in this case the release size, increases, the estimator converges to the true value with high probability (Bickell et. al. 1977). In most practical situations this estimator is reasonable. As an example of an impractical situation that leads to a ridiculous estimate, suppose that 10 fish are released, 20% of the catch is sampled, and 4 recoveries are observed. The method of moments estimate of the contribution rate is $\frac{4}{0.20 \cdot 10} = 2$.

An alternative estimator that does not lead to such potentially ridiculous estimates is the maximum likelihood estimator. mle. Numerically, however, the calculation of the mle is much more difficult since maximizing the likelihood involves summing over all possible values of tagged fish. a, in the catch. In particular the mle, $\hat{p}_{mle}$ is the value of p that will maximize

$$L(p; o, C, n) = \sum_{a=o}^{min(R,C-(n-o))} \binom{R}{a} p^a (1-p)^{R-a} \cdot \frac{\binom{a}{o}\binom{C-a}{n-o}}{\binom{C}{n}}.$$

Note that the likelihood function is expressed in terms of C and n, rather than f and b. b, the catch of other fish, is essentially a nuisance parameter, whose estimation can be bypassed since the total catch, C, is a fixed and known quantity after the fact. For large samples maximum likelihood estimators are approximately normally distributed with a mean equal to the parameter and a variance equal to the inverse of Fisher information (Bickell et. al. 1977). This calculation in practice would be very difficult.

Fortunately, given typical release sizes, the mle and the method of moments estimator probably do not differ very much. Since both are consistent estimators and the release sizes are generally very large, the method of moments estimator should be adequate.

A method of moments estimate of a contribution rate that is the sum of k sub-fishery contribution rates is simply the sum of the individual contribution rate estimates: $\hat{p}_{\cdot} = \sum_{i=1}^{k} \hat{p}_i$.

2.4   Variance and covariance estimation

For the sake of completeness we present the variance and covariances for contribution rate estimates. For some relatively simple hypotheses standardized test statistics will be approximately normally distributed and variance and covariance estimates are necessary. The testing procedures described later in the paper do not depend upon these estimates, however.

A variance estimate for $\hat{p}$ that was given in Newman (1990) is

$$V(\hat{p}) = p/R \cdot \left[q + \frac{(1-f)b}{f} \cdot \frac{(Rp+b)^2 - (1-p)}{(Rp+b)^3} \right]$$

An alternative variance estimate based on the assumption that total catch and

8

sample size are fixed numbers is

$$V(\hat{p}) = \frac{p}{f \cdot R} \left\{ 1 - pf + (R-1)p \left[\frac{n-1}{C-1} - f \right] \right\}.$$

Assuming that $C = pR + b$, the two estimators differ very slightly over a reasonable range of contribution rates.

Some useful extensions of these results include the variance for a contribution rate estimate that is the sum of several sub-fishery contribution rate estimates and the covariance of the contribution rate estimates from two different release groups. The results are summarized below and the derivations are in an appendix.

1. The variance for a contribution rate estimate from a single release group summed over several fisheries, say 1 through k, is

$$V(\sum_{i=1}^{k} \hat{p}_i) = \sum_{i=1}^{k} Var(\hat{p}_i) - 2\sum_{i<j}^{k} p_i p_j / R.$$

2. The covariance of the contribution rate estimates for a single fishery from two different release groups, say A and B, is

$$Cov(\hat{p}_A, \hat{p}_B) = p_A p_b \left[\frac{C(n-1)}{(C-1)n} - 1 \right]$$

3. Combining the above 2 results: the variance of the difference in contribution rates, summed over k fisheries, between two different release groups is

$$V(\sum_{i=1}^{k} \hat{p}_{Ai} - \sum_{i=1}^{k} \hat{p}_{Bi}) = \sum_{i=1}^{k} V(\hat{p}_{Ai}) + \sum_{i=1}^{k} V(\hat{p}_{Bi}) - \frac{2}{R_A} \sum_{i<j}^{k} p_{Ai} p_{Aj}$$
$$- \frac{2}{R_B} \sum_{i<j}^{k} p_{Bi} p_{Bj} - 2 \sum_{i=1}^{k} p_{Ai} p_{Bi} \left[\frac{C_i(n_i-1)}{(C_i-1)n_i} - 1 \right]$$

9

Note that for simple hypotheses of equality of contribution rate to a single fishery for two different release groups, these estimated variances could be used in the common normalized test statistic. I.e.,

$$z^* = \frac{\hat{\hat{p}}_A - \hat{\hat{p}}_B}{\hat{V}ar(\hat{\hat{p}}_A - \hat{\hat{p}}_B)}.$$

Over the range of contribution rates generally observed, rarely more than 15%, calculations of symmetric confidence intervals based on the above variance calculations and the standard normal approximation are not recommended. As pointed out in Schnute (1992) the exact confidence intervals based on the likelihood lead to asymmetric confidence intervals. It is the combined complexity of the likelihood function with compound distributions and the types of hypotheses to be tested in practice that motivated us to seek an alternative testing procedures. Before discussing the testing procedure, however, we briefly describe the two hypotheses commonly of interest to salmon fishery managers and scientists.

3 Hypotheses of interest

In this report we will focus on just two hypotheses, but the general procedure extends to testing more complex hypotheses. The first is that the contribution rates to a vector of r fisheries is the same for k different release groups. More succinctly,

$$H_o : \quad p_{11} = p_{21} = \ldots = p_{k1}$$
$$p_{12} = p_{22} = \ldots = p_{k2}$$
$$\vdots$$
$$p_{1r} = p_{2r} = \ldots = p_{kr}$$

10

The second hypothesis, often of interest to fishery managers, is that the catch profiles of two groups are the same. What we mean by catch profile is the proportion of fish caught by area (or time or both) conditional on being caught, e.g., 30% of the fish caught are caught off the Oregon coast, 50% off the Washington coast, and 20% off the British Columbia coast.

$$H_o: \quad r_{11} = r_{21} = \ldots = r_{k1}$$

$$r_{12} = r_{22} = \ldots = r_{k2}$$

$$\vdots$$

$$r_{1r} = r_{2r} = \ldots = r_{kr}$$

If there was no sampling of the catch, i.e., a complete census of the catch was taken, then standard statistical procedures for the comparison of multinomial samples could be conducted to test the first hypothesis. One approach would be a chi-square test of homogeneity (Snedechor et. al., 1980).

As mentioned earlier, in practice the single dimension outcome will often be the sum of contributions to several fisheries. Tests of equality of a 'summed' contribution rate do not necessarily imply that each individual component is equal. One could, for example, hypothesize that the combined contribution rate of some stock to fisheries A and B is 0.10 for release groups 1 and 2, but release group 1 contributes 0.05 to both fisheries while release group 2 contributes 0.03 to fishery A and 0.07 to fishery B. In most cases, however, we believe that the person formulating a hypothesis of equality is likely imagining that the two release groups are essentially identical in survival rates and migration pattern. Hence, equality at the combined level will generally be the result of equality at the component levels.

4 Monte Carlo and Bootstrap Testing Procedures

The iterated bootstrap test (Beran, 1988) is an extension of a bootstrap test and the bootstrap test is a variant of Monte Carlo testing (Besag et. al., 1977). The principle idea for all three methods is to simulate the distribution of the test statistic and estimate the p-value based on the simulated distribution.

Quite briefly, Monte Carlo tests are based on the rank order of a test statistic calculated from the collected data relative to computer generated test statistics based on the null hypothesis (Besag et. al. 1977, Ripley 1987). In other words, based on the parameter values specified by the null hypothesis, pseudo-random data is generated and the test statistic is calculated. This process is repeated several times and the rank order of the actual data's test statistic is calculated. If large test statistic values suggest deviations from the null hypothesis and the rank order of the test statistic is above the $(1-\alpha)$th percentile, then the null hypothesis can be rejected at the α level. The main advantage of Monte Carlo tests is that the sampling distribution of a test statistic does not have to be calculated, which can in some non-standard settings be quite difficult. A secondary advantage as stated by Besag (1977) 'the investigator is free to use a variety of informative statistics of his own choosing, rather than be dictated to by known theory'.

As an example of a Monte Carlo test, consider one of the simplest hypotheses-that a release group contributes at a specified rate to a specific fishery, i.e..

$$H_o : p = p_o.$$

Let $T = \frac{(\hat{p} - p_o)}{\sqrt{Var(\hat{p})}}$ serve as a test statistic. Deviations from the null hypothesis will be reflected in relatively large or small values of the test statistic. By relatively large or small is meant values that would occur only $\alpha\%$ of the time under the null

hypothesis. The procedure of Monte Carlo testing is to generate m-1 realizations of the test statistic under the null hypothesis and to use the rank order of the observed test statistic to provide an estimate of the p-value. If the observed test statistic's distribution does not depend on the value of the parameter, then this procedure provides an exact test at levels that are multiples of 1/m (Hall et. al. 1989).

For many of the more complicated hypotheses, such as

$$H_o : p_1 = p_2$$

or

$$H_o : p_1 = p_2 = \ldots = p_k,$$

the null hypothesis values are not specified. Therefore, the distribution of the data under the null hypothesis is not known. If a sample-based estimate of the parameters is substituted for the unknown parameter and the Monte Carlo approach is applied, one has a bootstrap test.

Bootstrap tests will in most cases be biased unless the test statistic's distribution under the null hypothesis does not depend on the true parameter value (Beran 1988). I.e., the probability that one is testing at will not equal the true probability of Type 1 errors (the probability of rejecting the null hypothesis when it is true). The magnitude of the bias, however, depends on the sample size, or with the compound distributions on release size (and to some degree on the catch sampling fraction).

Beran developed a procedure that provides a means of adjusting the bootstrap test p-value to get it closer to the true value (Beran 1988), whether or not the test statistic's null distribution depends on the unknown parameter value. There are some technical conditions that are sufficient for his results to hold. Primarily, one wants the estimated parameter value to be (root n) consistent, which the method

13

of moments estimates are, and that the bootstrap test statistic's distribution can be decomposed into an asymptotic expansion based on true test statistic's distribution. The latter condition we have not established, but simulations did not indicate any problems.

We refer to Beran's algorithm as the iterated bootstrap test. The procedure is as follows. Let the original sample size be j. Assume that a test statistic is chosen such that the larger its value, the less consistent the data is with the null hypothesis.

1. Using a sample based estimate of the unknown parameter(s), generate m realizations of the sample, i.e., parametric bootstrap samples of size j, and calculate the corresponding m test statistics. Let p_1 represent the relative rank or bootstrap p-value:

$$p_1 = \frac{(\#T^* \geq T) + 1}{m + 1}.$$

2. Now treating each of the m empirical bootstrap distributions as the true distributions, draw n samples of size j from each distribution (this is the 'iterated' bootstrap portion). Calculate n test statistics for each of the m bootstrap distributions. For each of the m sets, calculate the bootstrap p-values, say $p_i^*, i = 1 \ldots m$, using the first level bootstrap sample test statistic, $T_i^*, i = 1 \ldots m$ and the second level bootstrap sample test statistics. I.e., this is the first step repeated m times.

3. The relative rank of p_1 compared to the p_i^* sample is the adjusted or iterated bootstrap p-value, p_2.

$$p_2 = \frac{(\#p^* \leq p_1) + 1}{n + 1}.$$

If the adjusted p-value differs little from the original, first stage p-value, then the bias in the first stage bootstrap p-value is not too large.

4.1 Application of bootstrap tests to the compound distributions

To evaluate the testing procedure, we simulated recoveries of CWTs to fishery and calculated single and iterated bootstrap p-values. And as an application we examined a real data set based on recoveries of coho released from a coastal Washington hatchery.

For the simulation, comparisons were made between two release groups contributing the same fishery, i.e., a product binomial-multiple hypergeometric distribution. The null hypothesis is $H_o : p_1 = p_2$. The algorithm is as follows. For each iteration of the simulation two binomial random variables are generated. The catch of 'other' fish makes up the difference between the fixed total catch and the contributions from the two groups. A simple random sample (with size being a fixed fraction of the catch) is drawn without replacement and the observed recoveries of the three different groups are noted. Using the estimated contribution rate for a given iteration, 99 new realizations are simulated, the two test statistics calculated, and the bootstrap p-value determined. Within each of these 99 new realizations, the simulation is re-iterated to generate the iterated bootstrap p-value.

Adequacy of the procedure is evaluated graphically using histograms of the iterated bootstrap p-values and qq-plots against the uniform(0.1) distribution and analytically by Kolmogorov's goodness of fit test (Bickell et. al. 1977) applied to the uniform distribution. As an explanation of the reason for comparison to the uniform(0.1) distribution, consider the following. If the null hypothesis is true, and the test statistic is unbiased, then the probability of the test statistic exceeding the 1-α quantile of its distribution is simply α. So the observed p-value should be $\leq \alpha$, α% of the time- the definition of a uniform(0,1) random variable.

For the real data analysis, 18 release groups from the Grays Harbor region of

the Washington state coast are compared with respect to their contribution rates,

$$H_o : p_{1,j} = p_{2,j} = \ldots = p_{18,j}. \quad j = 1, \ldots, 11;$$

and contribution profile,

$$H_o : r_{1,j} = r_{2,j} = \ldots = r_{18,j}. \quad j = 1, \ldots, 11.$$

to 11 different fisheries. See tables 1 and 2 for details of releases and recoveries. *Table 1 n*

Under the null hypothesis of identical contribution rates, the combined estimate *here*
of contribution rates, a vector of length 12 (with the 12th category being fish dying *Table 2 n*
elsewhere), was used to generate the multinomial bootstrap samples. Within each *here*
catch a simple random sample was drawn without replacement thus generating
multiple hypergeometric random variables. The multiple hypergeometric samples
were drawn assuming that the sampling fraction was the same as that observed.
One point to make is that the fisheries were so highly aggregated spatially and
temporally the recoveries in actuality are drawn from samples of many catches.
Simulating the observed recoveries by drawing from the aggregated catches gener-
ates variability that is larger than what would be observed by drawing a stratified
random sample, which is somewhat closer, perhaps, to reality. The effect of larger
variability, however, is to make the testing procedure more conservative, i.e., less
likely to falsely reject the null hypothesis if in fact it is true.

To test the null hypothesis of identical contribution profiles, bootstrap samples
for each stock were generated separately using the stock-specific estimated contri-
bution rate to all 11 fisheries combined. Then for each stock the total contribution
was randomly assigned to each fishery using a common vector of proportions, es-
timated from all 18 stocks combined.

For both tests of the real data, n=99 and m=99 iterations were used, thus
generating 99 x 99 or 9.801 samples.

5 Results

Simulation results are shown in Figures 1 and 2 for contribution rates by two groups
of 0.01 to a single fishery. The release sizes were 10,000 for both groups, the total
catch was fixed at 2,000 fish (so the expected number of 'other' fish is 1,800),
and the sampling fraction was 0.20. Figures 1 and 2 are histograms and quantile-
quantile plots of the bootstrap and iterated bootstrap p-values, respectively, based
on 600 simulations. When the null hypothesis is true, then the p-values should
approximately follow the uniform(0,1) distribution. One can see from the plots
that the p-values follow the uniform distribution fairly closely. Kolmogorov's test
statistic was non-significant ($p \geq 0.05$).

Figure 1 ne here

Figure 2 ne here

For the Grays Harbor data, the hypothesis of equal contribution rates was easily
rejected, the p-value was 0.01 for both the single and iterated bootstraps. Con-
tribution rates to one of the British Columbia fisheries in particular varied widely
amongst the 18 stocks. On the other hand the hypothesis of equal contribution
profiles was not rejected, the iterated bootstrap p-value=0.68 (single bootstrap p-
value 0.63). So while the absolute contribution rates did differ significantly there
was no evidence suggesting that the contribution pattern differed significantly.

The rejection of the hypothesis of identical contribution rates for the 18 stocks
is not surprising when a plot of the rates is examined. Figure 3 is a plot of
the absolute contribution rates for all 18 stocks to the 11 fisheries. The line going
across the plot represents the estimated average contribution rate based on the null
hypothesis of equality. The contribution rates to the second fishery, which is the
British Columbia fishery 10, differs widely between stocks. The total contribution
rates to the fisheries differs widely as well.

Figure 3 ne here

On the other hand, the relative contribution rates, while still rather variable,
were consistent with the null hypothesis of identical contribution profiles. See

figure 4. Presumably the variation in relative contribution rates is high enough under the null hypothesis for the observed pattern to be well within what would be expected. While the survival rates differ between the stocks, the migration patterns (and vulnerability to fishing mortality) appear to be similar.

Figure 4 r.

here

6 Discussion

Building on the binomial-hypergeometric probability distribution, a computer intensive testing procedure was presented. Computer simulation-based comparisons of the procedures for one simple case demonstrated that both the single bootstrap and the iterated bootstrap procedures are accurate. The bootstrap test's primary advantage is its extreme flexibility, especially dealing with complex hypotheses. However, it can be very computer intensive, especially if the second iteration is thought necessary. Over the range of release sizes examined, both real and simulated, the difference between the bootstrap and iterated bootstrap p-values was minimal. This is likely due to the fact that the magnitude of the bias of the bootstrap test is quite small for release sizes on the order of 10,000. In practice, it may be quite acceptable to simply use a bootstrap test for the compound distribution for most test statistics of interest.

As an alternative to bootstrap tests, likelihood ratio test statistics can be calculated as well to test many hypotheses. The chief disadvantage, however, is the likelihood calculation quickly gets complicated for tests such as the test of equal contribution profiles. The likelihood ratio test statistic's null distribution would be approximated with the chi-square distribution and holds no real advantage over the bootstrap tests with respect to accuracy (Beran, 1988). An exact test procedure has been formulated for the special case of comparing two groups' contribution rates to a single fishery (J. Besag, Department of Statistics. University of Wash-

ington, Seattle, WA 98195, personal communication). This may be the focus of future efforts.

Another possibly important line to pursue is to study the power of various test statistics for particular hypotheses. Likelihood ratio statistics are most powerful in many settings, at least asymptotically, but difficult to construct for some hypotheses. A related issue is the use of 'studentized' test statistics, i.e., test statistics expressed in fractional for where the denominator is the standard error of the numerator. To some degree this can reduce the degree of dependence of the test statistic's distribution on the unknown parameter, and, therefore, improve the accuracy of the single stage bootstrap adequate. Simulations of compound distributions with moderately large release sizes, however, did not show noticeable differences in p-values between studentized and non-studentized statistics. Again the magnitude of the release sizes is such that the bias in single stage bootstrap tests is likely small even when the test statistics distribution depends on the unknown parameter.

Another approach is to approximate these numerically tedious compound distributions with analytically cleaner distributions. Simplifying approximations, such as binomial-binomial, say, or Poisson distributions have been suggested by Schnute (1992) and are implicit to generalized linear models for CWT recoveries (J. Schnute, Canadian DFO, Pacific Biological Station, Nanaimo, B. C. V9R 5K6, Canada, personal communication). However, based on analytical calculations of variance estimates for contribution rates, comparisons between binomial-hypergeometric, binomial-binomial. and Poisson distributions have yielded strikingly different results for many reasonable parameter values. The binomial-hypergeometric has a consistently lower variance than the other two procedures. The primary reason for the inefficiency of the approximations is due to the magnitude of the catch sampling fraction, ranging from 20 to 30%. The binomial approximation to the

19

hypergeometric works best when the sampling fraction is 5% or less. Still such approximations may be effective for simply detecting differences between groups.

Finally, there is a pressing data need. The compound distribution test statistics rely upon knowledge of the catch and sample size associated with each observed CWT recovery. Unfortunately many of the current data bases fail to provide this simple piece of information. Instead most CWT data bases contain information about sums of observed recoveries by a particular fishery, identified by time, area, and gear, and an estimate of actual recoveries, commonly called expanded recoveries. Further, the fishery is often the aggregation of several sub-fisheries, where sub-fisheries are defined as catches from which a single simple random sample is drawn. Since one component of variation in observed recoveries is clearly that induced by catch sampling, we do not believe that any testing procedure will be accurate that does not in someway account for that variation. The most straightforward answer to this problem is to ensure that data bases are detailed enough to provide catch and sample information at the level of each observed recovery.

Acknowledgements

References

Beran, R. 1988.. Prepivoting test statistics: a bootstrap view of asymptotic refinements. J. Am. Stat. Soc. **83**: 687-697.

Besag, J. and P. J. Diggle. 1977.. Simple monte carlo tests for spatial pattern. J.R. Statist. Soc. C. **83**: 327-333.

Bickell, P. J. and K. A. Doksum. 1977.. Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day, Inc./, San Francisco.

Clark, J. E. Clark and D. R. Bernard. 1987.. A compound multivariate binomial-hypergeometric distribution describing microwire tag recovery from commercial salmon catches in southeast Alasks. Alaska Department of Fish and Game (Juneau). Information Leaflet 261.

Hall, P. and D. M. Titterington. 1989.. The effect of simulation order on level accuracy and power of monte carlo tests. J. R. Statist. Soc. B. **51**:459-467.

Newman, K. B. 1990.. Variance estimation for contribution estimates based on sample recoveries of coded wire tagged fish. American Fisheries Society Symposium 7: 677-683.

Ripley, B. D. 1987.. Stochastic Simulation. John Wiley and Sons, New York, N. Y. 237 p.

Schnute, J. 1992.. Statistical analysis of embedded replicates in mark-recovery experiments. Can. J. Fish. Aquat. Sci. **49** (In press).

Snedechor, G. and W. Cochran. 1980.. Statistical Methods, 7th Ed. Iowa State Publishing, Ames, Iowa. 507 p.

A Derivation of variances and covariances for contribution rates

1. The variance for a contribution rate estimate from a single release group summed over several fisheries. say 1 through k, is

$$V(\sum_{i=1}^{k} \hat{p}_i) = \sum_{i=1}^{k} Var(\hat{p}_i) - 2\sum_{i<j}^{k} p_i p_j / R.$$

proof:

$$V(\sum_{i=1}^{k} \hat{p}_i) = \sum_{i=1}^{k} Var(\hat{p}_i) + 2\sum_{i<j}^{k} Cov(\hat{p}_i, \hat{p}_j)$$

where

$$
\begin{aligned}
Cov(\hat{p}_i, \hat{p}_j) &= E(\hat{p}_i \hat{p}_j) - p_i p_j \\
&= E_{a_i, a_j} \frac{1}{R^2} E(o_i/f_i \cdot o_j/f_j | a_i, a_j) - p_i p_j \\
&= \frac{1}{R^2} E_{a_i, a_j}(a_i \cdot a_j) - p_i p_j \\
&= \frac{R(R-1)p_i p_j}{R^2} - p_i p_j \\
&= -\frac{p_i p_j}{R}
\end{aligned}
$$

therefore

$$V(\sum_{i=1}^{k} \hat{p}_i) = \sum_{i=1}^{k} Var(\hat{p}_i) + 2\sum_{i<j}^{k} p_i p_j / R$$

$\square$

2. The covariance of the contribution rate estimates for a single fishery from two different release groups, say A and B, is

$$Cov(\hat{\hat{p}}_A, \hat{\hat{p}}_B) = p_A p_b \left[\frac{C(n-1)}{(C-1)n} - 1 \right]$$

proof:

$$
\begin{aligned}
Cov(\hat{\hat{p}}_A, \hat{\hat{p}}_B) &= E(\hat{\hat{p}}_A \hat{\hat{p}}_B) - p_A p_B \\
&= E_{a_A, a_B} E\left[\frac{C^2}{n^2 R_A R_B} o_A o_B | a_A, a_B \right] - p_A p_B \\
&= E_{a_A, a_B} \left[\frac{C^2}{n^2 R_A R_B} a_A a_B \frac{n(n-1)}{C(C-1)} \right] - p_A p_B \\
&= p_A p_B \frac{C(n-1)}{(C-1)n} - p_A p_B
\end{aligned}
$$

$\square$

Table 1. 18 Grays Harbor coho release groups' sizes and 11 ocean fisheries' catch and sample sizes. Releases are 1982 brood year and fisheries' catches are for 1985.

Release Data		Catch Data				
Tagcode	Release Size	Agency†	Yr	Fishery	Catch	Sample
211626	13,701	CDFO	85	23	681,358	182,978
632453	7,424	CDFO	85	10	2,241,052	445,504
632547	9,166	WDF	85	45	185,264	26,723
632743	48,935	WDF	85	42	83,163	40,965
632744	48,010	WDF	85	41	101,159	52,155
632745	50,538	WDF	85	23	1,101,170	228,554
632746	49,676	WDF	85	22	34,796	12,949
632861	45,404	WDF	85	15	79,443	9,640
632862	50,231	WDF	85	10	128,733	20,495
633017	48,976	ODFW	85	40	182,486	84,664
633018	49,913	ODFW	85	10	74,558	34,712
633027	31,602					
633046	10,760					
633047	20,910					
633048	18,055					
633061	10,717					
633062	7,663					
633107	4,415					

† CDFO=Canadian Department of Fisheries and Oceans, WDF=Washington Department of Fisheries, and ODFW=Oregon Department of Fisheries and Oceans.

Table 2. Detailed expanded recovery information for the 18 Grays Harbor coho release groups by tag code and fishery.

Tag Code:Fishery	CDFO 23	CDFO 10	WDF 45	WDF 42	WDF 41	WDF 23	WDF 22	WDF 15	WDF 10	ODFW 40	ODFW 10
211626	0	133	6	3	17	5	0	12	8	0	0
632453	0	6	0	0	0	0	0	0	0	0	4
632547	0	23	0	0	0	0	0	0	0	5	4
632743	0	56	0	8	0	0	19	7	0	0	0
632744	0	163	0	16	7	0	13	0	6	15	4
632745	4	128	2	4	17	0	29	0	0	6	0
632746	0	48	0	0	6	0	0	0	0	0	0
632861	0	172	0	4	6	10	21	12	6	3	2
632862	0	15	0	0	0	0	0	0	0	0	0
633017	0	31	0	4	5	0	0	5	0	9	3
633018	0	50	2	3	0	0	4	0	6	10	0
633027	0	87	2	0	7	0	9	0	6	0	0
633046	4	40	0	2	4	0	0	0	0	0	0
633047	0	112	0	0	11	0	8	19	11	3	0
633048	0	99	2	2	8	2	9	0	11	7	2
633061	0	21	0	2	0	0	2	0	0	0	0
633062	0	7	0	0	0	0	1	0	0	0	0
633107	0	10	0	3	5	0	3	0	0	2	0

Figure captions

- Fig. 1a. Histogram of bootstrap p-values based on 600 simulations.

- Fig. 1b. Histogram of iterated bootstrap p-values based on 600 simulations.

- Fig. 2a. Quantile-quantile plot of bootstrap p-values based on 600 simulations.

- Fig. 2b. Quantile-quantile plot of iterated bootstrap p-values based on 600 simulations.

- Fig. 3. Contribution rates to the 11 fisheries by the 18 groups. Line running through plot is the estimated average contribution rate for the 18 groups.

- Fig. 4. Relative contribution rates to the 11 fisheries by the 18 groups.

Fig. 1a. Histogram of Bootstrap p-values (600 simulations)
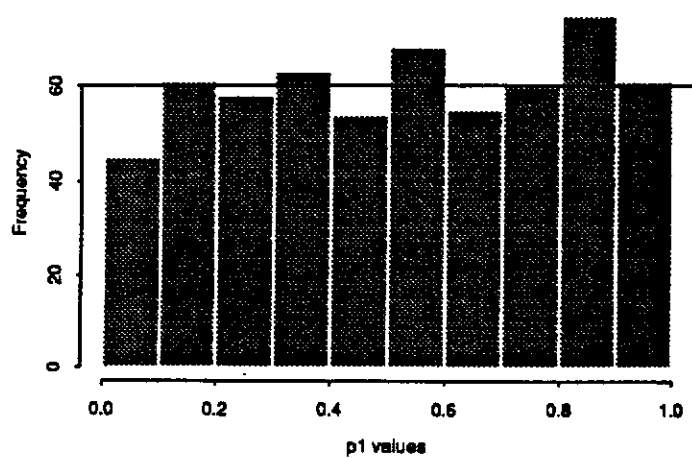
Fig. 1b. Histogram of Iterated Bootstrap p-values (600 simulations)
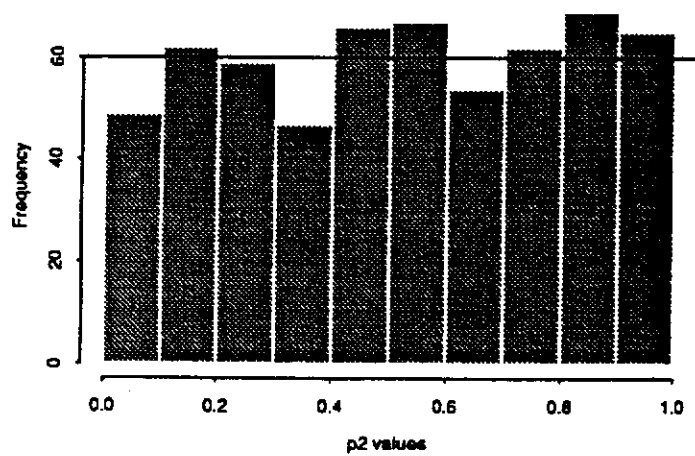
Fig. 2a QQ plot of Bootstrap p-values vs Uniform(0,1) distribution
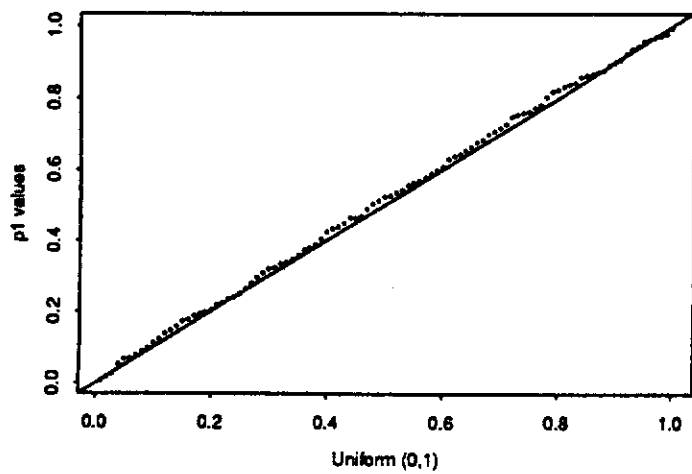

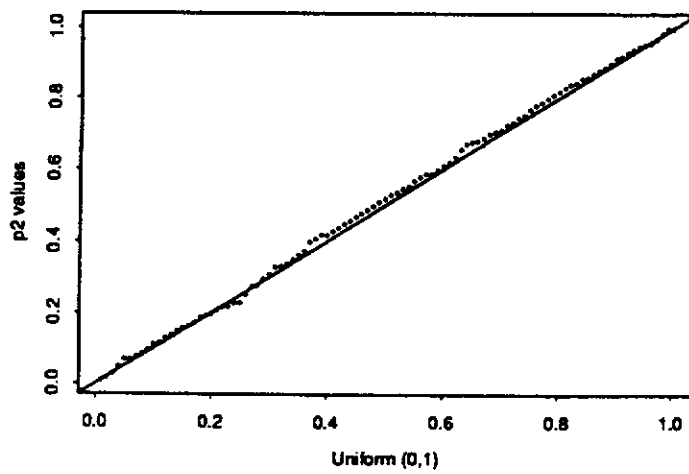Fig. 2b QQ plot of Iterated Bootstrap p-values vs Uniform(0,1) distribution

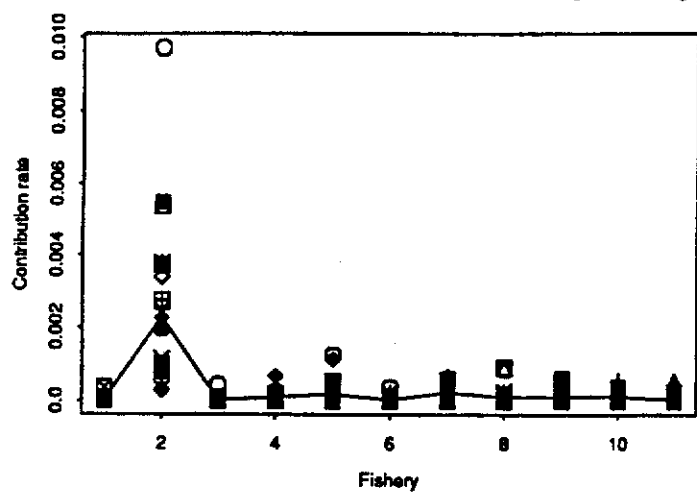Fig. 3. Contribution rates to the 11 fisheries by the 18 groups

Fig. 4. Relative contribution rates to the 11 fisheries by the 18 grou